

R E M A R K S

Claims 23-34 were rejected under 35 USC 112, first paragraph, for the same reason expressed by the Examiner in the previous Office Action. Although applicants respectfully disagree with the Examiner, to expedite prosecution the claims are amended to replace the word “tuple” with a phrase that means the same thing. As amended it is believed that the claims comply with 35 USC 112, first paragraph.

Claims 1-5, 7, 10, and 13-22 were rejected under 35 USC 103 as being unpatentable over Yang et al, US Patent 5,970,459 in view of Campbell et al, US Patent 6,366,883. Applicants respectfully traverse.

Claim 1:

The Examiner asserts that Yang et al teach the steps of inserting a plurality of phonemes, inserting duration specifications for the phonemes, and the step of “including at least one of said phonemes a time offset from the beginning of the duration of said phoneme that is greater than zero and less than the duration of said phone.” The Examiner also asserts that Yang et al teach “at least two prosody parameter specification toward a target value.” In support of the latter, the Examiner cites a passage in col. 4, lines 60-67 of Yang et al, which states:

The prosody processing unit 13 receives the processing results from the language processing unit 12, and calculates the values of the prosodic control parameters. The prosodic control parameter includes the time duration of phonemes, contour of pitch, contour of energy, position of pause, and length. The calculated results are transferred to the synchronization adjusting unit 15.

This passage teaches only about four prosodic control parameters; that is, (1) time duration of the phonemes, (2) contour of pitch, (3) contour of energy, and (4) position and length of pauses. What is taught about these four control parameters is that their values are calculated by the processing unit 12.

Thus, it is respectfully submitted that the above passage does not quite teach that which the Examiner asserts it teaches and, much more importantly, it does not teach the third step of claim 1.

The third step of claim 1 specifies

inserting, for at least one of said phonemes, a plurality of at least two prosody parameter specifications, with each specification of a prosody parameter specifying a target value for said prosody parameter, and a

point in time for reaching said target value, which point in time is unrestricted to any particular point within said duration, to thereby generate a signal adapted for converting into speech. (emphasis supplied)

The emphasized phrase in the above-quoted third step of the claim has three attributes: (a) a target value for said prosody, (b) a point in time for reaching the target value, and (c) the fact that the “time for reaching” is unrestricted. Focusing further, there are three words of importance in the claim: target, reaching, and time.

The word “target” is defined in the dictionary as “anything aimed or fired at.” Whatever eventually is **at** the target, it certainly is not at the target at an earlier time. Thus, one might have a bullet that starts not at a target but later reaches the target, a rocket that starts not at a target velocity but later reaches the target velocity, a signal that starts not at a target value but later reaches the target value, etc. Inherently, the notion is of not being at some goal, taking action to reach the goal, and eventually (if successful) reaching that goal after some elapsed time.

Thus, the three words of importance in the claim (target, reaching, and time) are part of one cohesive notion. What the claim specifies is that a prosody parameter is **specified with** two attributes that encompass the three words of importance: a target value, and a time for reaching the target value.

With this understanding in mind, the following table compares the teaching of the passage cited by the Examiner to the claim language.

Reference	Relevance to the third step of the claim
<u>time duration of phonemes</u>	Time duration is addressed in the <u>second</u> step of claim 1 and not to the third step and, therefore, has no relevance to the third step of the claim.
<u>contour of pitch:</u> The dictionary defines the word “contour” as “the outline of a figure, body, or mass, or a line that represents such an outline.” To place the reference’s teaching on the best possible footing in view of this definition, one might say that processor 12 computes a value that elects, or identifies, a contour of pitch, and the selected contour specifies pitch values	Clearly a computed value that selects a contour is NOT a target that is aimed at, and hence is not a “target value,” as the claim specifies. Moreover, the claim imposes the limitation that the target value is reached at a specified point in time. No specification of a <u>time to reach</u> is found in the value that is computed by language unit 12.

as they change from one instant of time to the next.	
<u>contour of energy</u> : The observations above hold. That is, to place the reference's teaching on the best possible footing, one might say that processor 12 computes a value that elects, or identifies, a contour of energy, and the selected contour specifies energy values as they change from one instant of time to the next.	The argument above is applicable here as well. That is, there is no teaching of a target value, and there is no teaching of a specification as to when the target value is reached.
position and length of pause:	The nature of a target by means of some path to reach the target inherently includes the notion that one can be far from the target, close to the target, or hit/reach the target. In contradistinction, either there is a pause, or there isn't one. It's a binary condition. Therefore, in the case of a pause attribute, there is no "target value," and there is no "time to reach" the target.

Thus, a perusal of the above comparison table reveals that the reference fails to teach the third step of claim 1.

Indeed, the Examiner admits that Yang et al fail to "explicitly teach any selected point in time for reaching said target value," but points to a passage at col. 16, lines 14- col. 17, line 23 of the Campbell et al reference that allegedly teaches "a selected point in time for reaching the target value."

Applicants respectfully disagree on two grounds. First, even if the Campbell et al reference did teach that which the Examiner asserts, it would still not be relevant because, as discussed above, the notion of having a target value that is reached at a specified point in time within the duration of a phoneme is not applicable to a "contour of pitch" parameter value and is not applicable to a "contour of energy" parameter value, and the notion of having a target value to be reached is not applicable for pauses.

Put another way, even if Campbell et al were to "explicitly teach any selected point in time for reaching said target value," a skilled artisan would not specify a target and a point in time for reaching the target for a "contour of pitch" parameter, a "contour of energy" parameter, or a "position and length of pause" parameter, because there is absolutely no reason for an artisan who specifies an entire contour of pitch by specifying

one value, to modify his/her design to specify the pitch at one point in time by specifying **two values**.

Second, Campbell et al do not teach the notion of a parameter having a target value that is to be reached at a specified point in time. The passage cited by the Examiner consists of a number of paragraph, and because of this the following table quotes each paragraph in a separate left-column cell, and applicants' comments about the paragraph are to the right of the cell:

<p>At step S14, the start position and end position in the speech waveform database file composed of either a plurality of sentences or one sentence for each phoneme segment are recorded, and an index number is assigned to the file. Next, at step S15, the first acoustic feature parameters for each phoneme segment are extracted by using, for example, a known pitch extraction method. Then, at step S16, the phoneme labeling is executed for each phoneme segment, and the phoneme labels and the first acoustic feature parameters for the phoneme labels are recorded. Further, at step S17, the first acoustic feature parameters for each phoneme segment, the phoneme labels and the first prosodic feature parameters for the phoneme labels are stored in the feature parameter memory 30 together with the file index number and the start position and time duration in the file. Finally, at step S18, index information including the index number of the file and the start position and time duration in the file are given to each phoneme segment, and the index information is stored in the feature parameter memory 30, then the speech analysis process is completed.</p>	<p>This paragraph speaks of feature parameters, but it does NOT speak of values of the parameters (targets or otherwise). The only reference to duration is in the last sentence where it is mentioned that the start positions and durations of <u>phonemes</u> are given. The start positions of phonemes are understood to be the start positions of the phonemes in the sequence of phonemes that combine to form an utterance (e.g. a sentence).</p>
<p>FIGS. 5 and 6 are flowcharts of the weighting coefficient training process which is executed by the weighting coefficient training controller of FIG. 1.</p>	<p>This paragraph says nothing of parameters, values, or times.</p>
<p>Referring to FIG. 5, first of all, at step S21, one phonemic kind is selected from the feature parameter memory 30. Next, at step S22, the second acoustic feature parameters are extracted from the first acoustic feature parameters of a phoneme that has the same phonemic kind as the selected phonemic kind, and then, are taken as the second acoustic feature parameters of the target phoneme. <i>Then, at step S23, the Euclidean cepstral distances of acoustic distances between the remaining phonemes other than the target phoneme that have the same phonemic kind, and the target</i></p>	<p>The notion of a "target" is found in the highlighted (gray) sentence, but addresses target <u>phoneme</u>, not a target value of a parameter of a phoneme. The notion of a target phoneme is understood to mean the phoneme that one ought to select.</p>

<p><i>phoneme in the second acoustic feature parameters, as well as the log phoneme duration with the bottom of 2 are calculated.</i> At step S24, it is decided whether or not the processes of steps S22 and S23 have been done on all the remaining phonemes. At step S24, if the processes have not been completed for all the remaining phonemes, another remaining phoneme is selected at step S25, and then, the processes of step S23 and the following thereto are iterated.</p>	
<p>On the other hand, if the processing has been completed at step S24, the top N1 best phoneme candidates are selected at step S26 based on the distances and time durations obtained at step S23. Subsequently, at step S27, the selected N1 best phoneme candidates are ranked into the first to N1-th places. Then, at step S28, for the ranked N1 best phoneme candidates, the scale conversion values are calculated by subtracting intermediate values from the respective distances. Further, at step S29, it is decided whether or not the processes of steps S22 to S28 has been completed for all the phonemic kinds and phonemes. If the processes of steps S22 to S28 have not been completed for all the phonemic kinds, another phonemic kind and phoneme is selected at step S30, and then the processes of step S22, and the following are iterated. On the other hand, if the processes of steps S22 to S28 has been completed for all the phonemic kinds at step S29, the program flow goes to step S31 of FIG. 6.</p>	<p>This paragraph discusses selecting “best phoneme candidates” based on distances and time durations. This has nothing to do with target values of a <u>phoneme parameter</u>, and it has nothing to do with time to reach the target values.</p>
<p>Referring to FIG. 6, at step S31, one phonemic kind is selected. Subsequently, at step S32, the second acoustic feature parameters for each phoneme are extracted for the selected phonemic kind. Then, at step S33, by performing the linear regression analysis based on the scale conversion value for the selected phonemic kind, the degrees of contribution to the scale conversion values in the second acoustic feature parameters are calculated, and the calculated degrees of contribution are stored in the weighting coefficient vector memory 31 as weighting coefficients for each target phoneme. At step S34, it is decided whether or not the processes of steps S32 and S33 has been completed for all the phonemic kinds. If the processes have not been completed for all the phonemic kinds at step S34, another phonemic kind is selected at step S35, and the processes of step S32 and the following are iterated. On the other hand, if the processes has been completed for all the phonemic kinds at step S34, the weighting coefficient training process is completed.</p>	<p>This paragraph speaks of extracting feature parameters and performing regression analysis. There is no mention here of target values, or of times for reaching these target values.</p>

Thus, it is respectfully submitted that the passage cited by the Examiner does NOT teach “a selected point in time for reaches the target value” of anything.

In view of the above arguments, it is respectfully submitted that the third step of claim 1 is not obvious in view of Yang et al and Campbell et al combination of references.

Remaining Claims:

Claims 2-5, 7 and 10-20 depend on claim 1 and therefore are believed to not be obvious in view of the Yang et al and Campbell et al combination of references at least by virtue of this dependence. Additionally, it is believed that at least some of the claims contain limitations that make the claims patentable over the Yang et al and Campbell et al combination of references.

Amended claim 2 specifies that at least one phoneme has a specification that includes at least two parameter specifications that BOTH specify pitch. No such notion exists in Yang et al, in col. 7, line 65 (cited by the Examiner) or elsewhere in the reference.

Claim 7 specifies the time at which a parameter reaches its target more particularly, and as indicated above, the entire notion of a parameter reaching a target (i.e. starting at some value other than the target, and traversing some path that eventually makes the parameter have the target value at a given time) is simply not present in either the Yang et al reference or the Campbell et al reference.

Although independent claim 21 was rejected in the group of claims identified in item 4 of the Office action, no explicit comments are offered by the Examiner to justify the rejection. Amended claim 21 is believed not obvious in view of the Yang et al and the Campbell et al combination of references for the reasons set forth above. Additionally, amended claim 21 explicitly limits the claim to specifications where at least one phoneme has at least one specification that consists of a target value, a time offset, and a delimiter therebetween. No notion of such form to the specification of a control parameter is found in or suggested by either of the cited references.

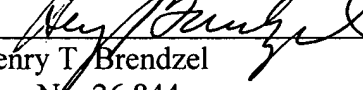
As for claim 22, the Examiner has also not provided an explicit comment that explains the reason for the rejection. Applicants note, however, that claim 22 is dependent on claim 21, which is believed to not be obvious in view of the cited

references. Moreover, there is no teaching anywhere in the references that the value of a parameter is not restricted, except at the specified offset time. In contradistinction, the contour of pitch and the contour of energy are clearly specified throughout the phoneme's duration (by virtue of the definition of a "contour"), and a pause is also clearly defined throughout – i.e., it being a pause. Therefore, it is respectfully submitted that claim 22 is not obvious in view of the Yang et al and the Campbell et al combination of references.

In light of the above amendments and remarks, applicants respectfully submit that all of the Examiner's rejections have been overcome. Reconsideration and allowance are respectfully solicited.

Respectfully,
Mark Beutnagel
Joern Ostermann
Schuyler Quackenbusch

Dated: 11/17/04

By 
Henry T. Brendzel
Reg. No. 26,844
Phone (973) 467-2025
Fax (973) 467-6589
email brendzel@comcast.net